AMENDMENT

This listing of claims will replace all prior versions, and listings, of claims in the application:

Listing of Claims:

1. (Currently Amended) A method for segmenting multi-speaker speech data by speaker, the method comprising:

detecting speaker changes in multi-speaker speech data to obtain an initial segmentation of the multi-speaker speech data, wherein estimated segments are generated by the detected speaker changes;

clustering the estimated segments into groups of estimated segments, wherein each group of estimated segments is associated with a single speaker;

checking whether segments in a first group of estimated segments overlap segments in a second group of estimated segments, wherein if segments of the first group overlap with segments of the second group, then the method comprises pooling the first and second group; and

modeling and resegmenting ~~each group~~ any pooled groups and remaining groups of estimated segments to obtain stable segmentations~~; and~~

~~checking overlap between segments in each group of estimated segments, wherein each group of estimated segments is associated with a different speaker when the overlap is below a specified threshold~~.

2. (Original) A method as defined in claim 1, wherein detecting speaker changes in multi-speaker speech data to obtain an initial segmentation of the multi-speaker speech data further comprises performing a front-end analysis on the multi-speaker speech data.

7. (Original) A method as defined in claim 1, wherein clustering the estimated segments into groups of segments further comprises applying an agglomerative hierarchical clustering procedure to obtain an initial grouping of segments.

8. (Original) A method as defined in claim 1, wherein clustering the estimated segments into groups of estimated segments further comprises clustering the estimated segments into groups of estimated segments until all of the estimated segments are merged into a final group, wherein the final group includes one or more clusters that correspond to the groups of estimated segments and wherein each cluster corresponds to a single speaker.

9. (Previously Presented) A method as defined in claim 8, further comprising identifying the one or more clusters in the final group empirically.

10. (Original) A method as defined in claim 1, wherein each estimated segment is initially in a separate group of estimated segments, wherein clustering the estimated segments into groups of estimated segments further comprises:

modeling each estimated segment by a low-order Gaussian mixture model;

generating table of pairwise distances using the low-order Gaussian mixture models, wherein the table of pairwise distances includes a distance between each estimated segment and every other estimated segment; and

merging at least two groups of estimated segments to produce a new group of estimated segments such that a merger of the at least two groups of estimated segments produces a smallest increase in the distance.

7. (Original) A method as defined in claim 1, wherein clustering the estimated segments into groups of segments further comprises applying an agglomerative hierarchical clustering procedure to obtain an initial grouping of segments.

8. (Original) A method as defined in claim 1, wherein clustering the estimated segments into groups of estimated segments further comprises clustering the estimated segments into groups of estimated segments until all of the estimated segments are merged into a final group, wherein the final group includes one or more clusters that correspond to the groups of estimated segments and wherein each cluster corresponds to a single speaker.

9. (Previously Presented) A method as defined in claim 8, further comprising identifying the one or more clusters in the final group empirically.

10. (Original) A method as defined in claim 1, wherein each estimated segment is initially in a separate group of estimated segments, wherein clustering the estimated segments into groups of estimated segments further comprises:

modeling each estimated segment by a low-order Gaussian mixture model;

generating table of pairwise distances using the low-order Gaussian mixture models, wherein the table of pairwise distances includes a distance between each estimated segment and every other estimated segment; and

merging at least two groups of estimated segments to produce a new group of estimated segments such that a merger of the at least two groups of estimated segments produces a smallest increase in the distance.

11. (Original) A method as defined in claim 10, further comprising merging new groups of estimated segments until all estimated segments are merged into a final group.

12. (Original) A method as defined in claim 10, wherein modeling and resegmenting each group of estimated segments to obtain stable segmentations further comprises:

constructing a Gaussian mixture model for each group of estimated segments; and

calculating a frame-by-frame likelihood ratio detection score for each Gaussian mixture model compared with a Gaussian mixture model representing the multi-speaker speech data.

13. (Original) A method as defined in claim 1, wherein checking overlap between segments in each group of estimated segments further comprises:

pooling the estimated segments that overlap; and

modeling and resegmenting the estimated segments that overlap.

14. (Original) A method as defined in claim 1, further comprising performing post-processing on the speaker segments by creating a segmentation lattice, wherein a best path through the segmentation lattice is a sequence of non-overlapping estimated segments such that an overall segmentation likelihood is maximized.

15. (Original) A method as defined in claim 1, further comprising obtaining a final segmentation by:

comparing detection scores of each group of estimated segments;

hypothesizing segment boundaries when a difference between detection scores crosses zero; and

accepting segments defined by the hypothesized segment boundaries if each segment has a duration above a duration threshold and if each segment does not cross a silence gap that is longer than a gap threshold.

16. (Original) A method as defined in claim 1, wherein the speech data is one of a telephone conversation between two or more speakers; an archived recorded broadcast news program; and a recorded meeting between multiple speakers.

17. (Currently Amended) A method for segmenting speech data into speaker segments by speaker, the method comprising:

scanning input speech data with a windowed generalized likelihood ratio (GLR) function to obtain speech segments, wherein the input speech data includes a plurality of speakers;

clustering the speech segments into one or more clusters, wherein each cluster is associated with a single speaker;

if more clusters exist than speakers, then:

checking overlap between segments in each cluster;

pooling clusters that have overlap between at least one segment in each pooled cluster; and

resegmenting and remodeling the pooled clusters;

creating models for each cluster; and

rescanning the input speech data with the models to resegment the speech data and obtain speech segments for each speaker included in the speech data.

hypothesizing segment boundaries when a difference between detection scores crosses zero; and

accepting segments defined by the hypothesized segment boundaries if each segment has a duration above a duration threshold and if each segment does not cross a silence gap that is longer than a gap threshold.

16. (Original) A method as defined in claim 1, wherein the speech data is one of a telephone conversation between two or more speakers; an archived recorded broadcast news program; and a recorded meeting between multiple speakers.

17. (Currently Amended) A method for segmenting speech data into speaker segments by speaker, the method comprising:

scanning input speech data with a windowed generalized likelihood ratio (GLR) function to obtain speech segments, wherein the input speech data includes a plurality of speakers;

clustering the speech segments into one or more clusters, wherein each cluster is associated with a single speaker;

if more clusters exist than speakers, then:

checking overlap between segments in each cluster;

pooling clusters that have overlap between at least one segment in each pooled cluster; and

resegmenting and remodeling the pooled clusters;

creating models for each cluster; and

rescanning the input speech data with the models to resegment the speech data and obtain speech segments for each speaker included in the speech data.

3. (Original) A method as defined in claim 1, wherein detecting speaker changes in multi-speaker speech data to obtain an initial segmentation of the multi-speaker speech data further comprises at least one of:

detecting speaker changes using Bayes Information Criterion; and

detecting speaker changes using a generalized likelihood ratio formulation, wherein a speaker change occurs when the generalized likelihood ratio formulation exhibits a dip.

4. (Original) A method as defined in claim 1, wherein detecting speaker changes in multi-speaker speech data to obtain an initial segmentation of the multi-speaker speech data further comprises estimating speaker segments by detecting dips in the generalized likelihood ratio formulation.

5. (Original) A method as defined in claim 1, wherein detecting speaker changes in multi-speaker speech data to obtain an initial segmentation of the multi-speaker speech data further comprises estimating speaker segments when the generalized likelihood ratio formulation remains above a specified threshold for a particular duration.

6. (Original) A method as defined in claim 1, wherein detecting speaker changes in multi-speaker speech data to obtain an initial segmentation of the multi-speaker speech data further comprises determining a location of a boundary between speaker segments by calculating the generalized likelihood ratio formulation over successive overlapping windows throughout the multi-speaker speech data.

7. (Original) A method as defined in claim 1, wherein clustering the estimated segments into groups of segments further comprises applying an agglomerative hierarchical clustering procedure to obtain an initial grouping of segments.

8. (Original) A method as defined in claim 1, wherein clustering the estimated segments into groups of estimated segments further comprises clustering the estimated segments into groups of estimated segments until all of the estimated segments are merged into a final group, wherein the final group includes one or more clusters that correspond to the groups of estimated segments and wherein each cluster corresponds to a single speaker.

9. (Previously Presented) A method as defined in claim 8, further comprising identifying the one or more clusters in the final group empirically.

10. (Original) A method as defined in claim 1, wherein each estimated segment is initially in a separate group of estimated segments, wherein clustering the estimated segments into groups of estimated segments further comprises:

modeling each estimated segment by a low-order Gaussian mixture model;

generating table of pairwise distances using the low-order Gaussian mixture models, wherein the table of pairwise distances includes a distance between each estimated segment and every other estimated segment; and

merging at least two groups of estimated segments to produce a new group of estimated segments such that a merger of the at least two groups of estimated segments produces a smallest increase in the distance.

11. (Original) A method as defined in claim 10, further comprising merging new groups of estimated segments until all estimated segments are merged into a final group.

12. (Original) A method as defined in claim 10, wherein modeling and resegmenting each group of estimated segments to obtain stable segmentations further comprises:

constructing a Gaussian mixture model for each group of estimated segments; and

calculating a frame-by-frame likelihood ratio detection score for each Gaussian mixture model compared with a Gaussian mixture model representing the multi-speaker speech data.

13. (Original) A method as defined in claim 1, wherein checking overlap between segments in each group of estimated segments further comprises:

pooling the estimated segments that overlap; and

modeling and resegmenting the estimated segments that overlap.

14. (Original) A method as defined in claim 1, further comprising performing post-processing on the speaker segments by creating a segmentation lattice, wherein a best path through the segmentation lattice is a sequence of non-overlapping estimated segments such that an overall segmentation likelihood is maximized.

15. (Original) A method as defined in claim 1, further comprising obtaining a final segmentation by:

comparing detection scores of each group of estimated segments;

hypothesizing segment boundaries when a difference between detection scores crosses zero; and

accepting segments defined by the hypothesized segment boundaries if each segment has a duration above a duration threshold and if each segment does not cross a silence gap that is longer than a gap threshold.

16. (Original) A method as defined in claim 1, wherein the speech data is one of a telephone conversation between two or more speakers; an archived recorded broadcast news program; and a recorded meeting between multiple speakers.

17. (Currently Amended) A method for segmenting speech data into speaker segments by speaker, the method comprising:

scanning input speech data with a windowed generalized likelihood ratio (GLR) function to obtain speech segments, wherein the input speech data includes a plurality of speakers;

clustering the speech segments into one or more clusters, wherein each cluster is associated with a single speaker;

if more clusters exist than speakers, then:

checking overlap between segments in each cluster;

pooling clusters that have overlap between at least one segment in each pooled cluster; and

resegmenting and remodeling the pooled clusters;

creating models for each cluster; and

rescanning the input speech data with the models to resegment the speech data and obtain speech segments for each speaker included in the speech data.

18. (Original) A method as defined in claim 17, wherein scanning input speech data with a windowed GLR function to obtain speech segments further comprises performing a front-end analysis on the input speech sample.

19. (Original) A method as defined in claim 17, wherein scanning input speech data with a windowed GLR function to obtain speech segments further comprises:

deriving a Gaussian Mixture Model for each window of the windowed GLR function from a Gaussian Mixture Model of the input speech data; and

adapting component weights in each window.

20. (Original) A method as defined in claim 19, wherein each window generates stable statistics and only includes one speaker segment change.

21. (Original) A method as defined in claim 17, wherein scanning input speech data with a windowed GLR function to obtain speech segments further comprises estimating speech segments by detecting dips in the windowed GLR function.

22. (Original) A method as defined in claim 17, wherein scanning input speech data with a windowed GLR function to obtain speech segments further comprises estimating speech segments when the windowed GLR function remains above a specified threshold for a particular duration.

23. (Original) A method as defined in claim 17, wherein clustering the speech segments into one or more clusters further comprises obtaining an initial grouping of speech segments using an agglomerative hierarchical clustering procedure.

24. (Original) A method as defined in claim 23, wherein obtaining an initial grouping of speech segments using an agglomerative hierarchical clustering procedure further comprises:

generating a table of pairwise distances that defines a distance between each speech segment and every other speech segment; and

merging estimated segments to form groups of speech segments, wherein each merger produces a smallest increase in distance between speech segments included in each group of speech segments.

25. - 26. (Cancelled)

27. (Original) A method as defined in claim 17, further comprising performing post-processing on the speaker segments, wherein performing post-processing on the speaker segments further comprises creating a segmentation lattice, wherein a best path through the segmentation lattice is a sequence of non-overlapping speaker segments.

28. (Original) A method as defined in claim 17, further comprising comparing speaker segments to a set of known target speakers to detect, label and locate their presence in speech data.

29. (Original) A method as defined in claim 17, further comprising:

comparing detection scores of each group of estimated segments;

hypothesizing segment boundaries when a difference between detection scores crosses zero; and

accepting segments defined by the hypothesized segment boundaries if each segment has a duration above a duration threshold and if each segment does not cross a silence gap that is longer than a gap threshold.

30. (Original) A method as defined in claim 17, wherein the input speech data is one of a telephone conversation between two or more speakers; an archived recorded broadcast news program; and a recorded meeting between multiple speakers.

31. (Currently Amended) A method for segmenting speech data by speaker, the method comprising:

obtaining initial estimated segments of the speech data, wherein the estimated segments are unlabeled;

clustering the initial estimated segments until the initial estimated segments are grouped into a final group;

selecting one or more clusters from the final group, wherein one or more clusters corresponds to groups of estimated segments;

iteratively modeling and resegmenting each group of estimated segments until changes in segment boundaries for the estimated segments in each group of estimated segments from a particular iteration to a next iteration are below a threshold;